Genetic Anomaly Based Ids

[#] M. Jagadheeswari, *Dr. M. Anand Kumar
[#]*Ph.D. Research Scholar, Dept. of Computer Science, Karpagam University, Tamil Nadu, India*
**Associate Professor, Dept. of Information Technology, Karpagam University, Tamil Nadu, India*

Abstract

The security of network devices will be great issues to provide quality of network. Intrusion detection system have been used many techniques to identify, detect and classify attacks that have been proposed, developed and tested either in offline or online mode. Clustering based detection technique is used to find out the dissimilarity measure to form the k clusters. It represents genetic process specified each chromosome of centroids of the clusters. Two stage fitness function proposed: i) refine the clustering function to introduce the confidence interval ii) calculate and maximize the inter-cluster variance

Keywords

anomaly based IDS, Genetic algorithm, Clustering.

I. INTRODUCTION

Intrusion means most set of actions that compromise the integrity, confidentiality or availability of a resource. An IDS is set of components and its associated function to monitor network function and host activity to detect and react attack intrusion. Based on detection technique, IDS classified into two techniques i) Anomaly detection and ii) Misuse detection. Misuse behavior means list of signature of known attack. If any attack is match with preexisting signature, attack states in the network. Anomaly detection means that behavior pattern deviates from the normal behavior. If the deviation between observation and normal behavior is exceed the predefined threshold value, and then alarm will be raised. Signature based IDS provide good detection rate from predefined database signature. Its main disadvantage is cannot able to find unknown pattern. But anomaly detection can easily find out unknown attack and increase false positive rate.

II. CLUSTER ANALYSIS

Cluster analysis find out the groups in the data. The main three partition method is used in anomaly based IDS: hierarchical approach, non-hierarchical approach and biomimetic approach. Hierarchical approach builds hierarchy of cluster using the bottom-up method and every observation starts with its own cluster. Each node merged with upper node in hierarchy order. In non-hierarchy approach, set the random initialization and termination condition. Biomimetic approach works based on the human biology.

Anomaly based cluster techniques used two types of fitness function:

i) Confidence interval to redefine the cluster
ii) Maximize the covariance of the matrix.

The main steps of clustering algorithms are:

a) Total number of clusters, cluster centroids.
b) Distance and density between clusters.
c) Fixed configuration and iteration number.

Partitioning approach specified into two processes.

i) Initiate partitioning solution.
ii) Optimizing partitioning.

A. *Genetic algorithm*

A genetic algorithm is an robust search algorithm that starts by initializing solutions encoded into string chromosome. The format of the string may be string, number or bits based on its attributes. Each and every chromosome has a threshold value to measure its goodness. Each and every sample will be evaluated and assigning the fitness value, apply the genetic operator selection, crossover and mutation. This action will be continuing until it met the termination condition.

B. *Clustering And Genetic Steps*

Genetic population closely related to solve the problem and it is a key element to the genetic processes. It follows four steps to proceed.

i) String codification:

String codification represents types of data to be stored into the string. It may be binary string, floating point, character string or other types of data structure.

ii) Initial clustering:

Each chromosome can be represented in discriminate distance. To verify the proposed method, calculated Euclidean distance. Each chromosome assign a thread and calculate the fitness function.

iii) fitness function:

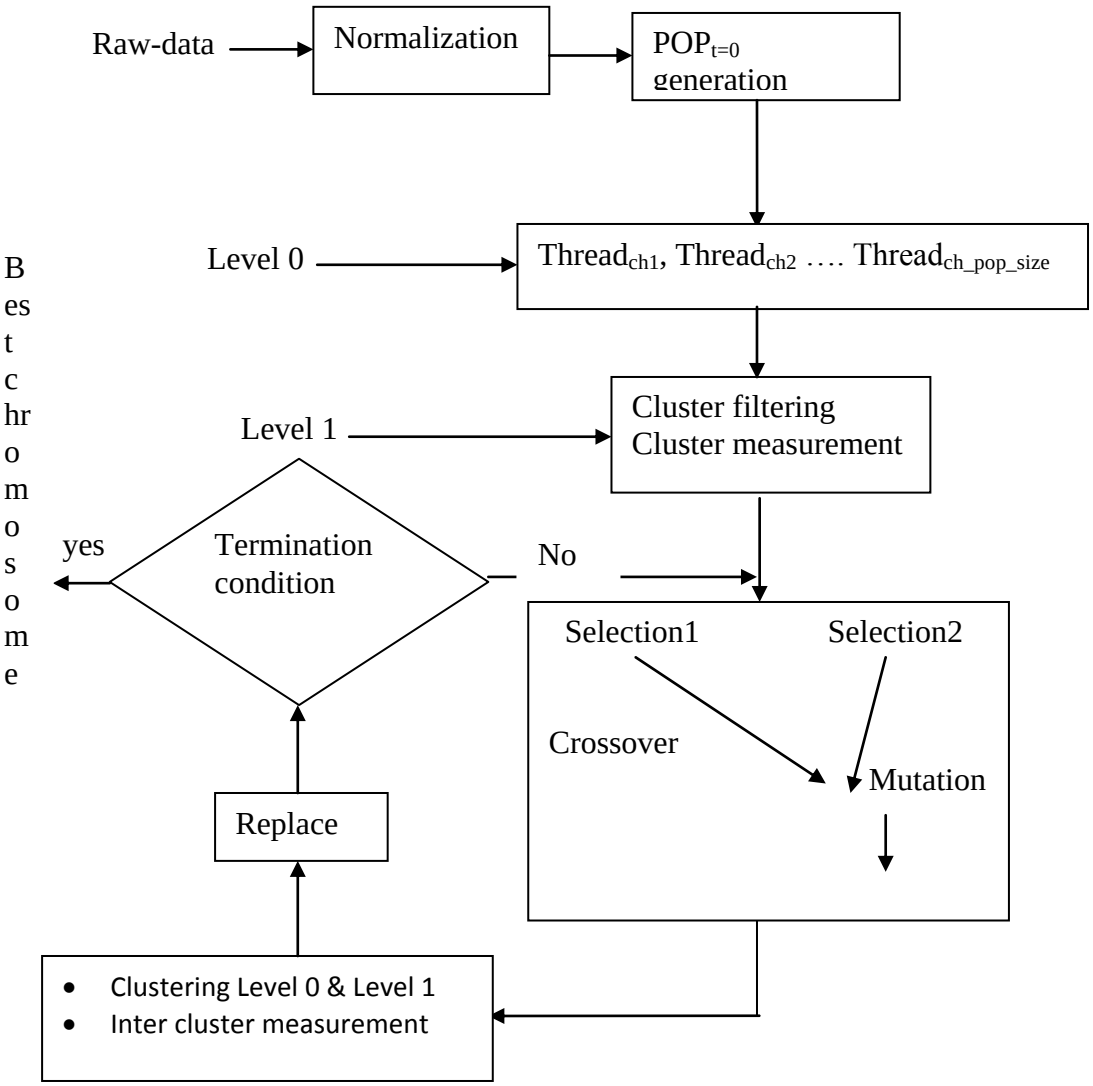
The fitness function means calculate the goodness of each chromosome. Closest

chromosome for the optimal solution finds out the highest fitness value. It will ensure the good individual will retain for the next generations. If less suited one will be discarded.

IV. GENETIC OPERATOR

A. *Selection:*

Selection process randomly picks two pair of chromosome based on the higher fitness function.



Fitness function will be evaluates in two levels:

- intrinsic and
- Extrinsic.

a) Intrinsic level:

Filtering the cluster represents in confidential interval and it can be reallocated to clusters represents the center by the chromosome. Through this requirement, rejection cluster will be created. The size of dataset will be depend on its selected attributes, normal instance frequency and anomaly instance.

b) Extrinsic level:

Chromosomes quality will be assessed by better quality of its clusters.

It is systematically taking best chromosomes with better fitness value. It promotes the best chromosome to be selected and give less chance of individuals to be selected. It will use three types of selection method: A) Roulette wheel selection B) Rank selection C) Tournament selection. Tournament selection means randomly select a number from the cluster and iterate to the next generation. It is very efficient that leads to an optimal solution.

B. *Crossover*

Crossover is a parent solution and derived a child from cluster. Genetic algorithm uses a crossover operator to achieve a large jump. Parents should be swapped to form a new cluster.

C) Mutation:

Crossover operation will be less efficient over individual time. Mutation will become more important. It will create a better solution to the problem, it will make a small jump to avoid uniform population.

Simulated attack represents through four categories:

- DOS(Denial Of Service) attack aim to make a resource unavailable.
- U2R (User to Root) attack tries to obtain administrator privileges.
- R2L (Remote to Local) attack attempt to access the remote machine access through network.
- PROBE represents collect information from user or security policy.

V. CONCLUSION

Anomaly based detection scheme used unsupervised mode using genetic process. Clustering genetic anomaly based IDS obtain normal and anomaly homogenous partitioning. Cluster instance will be rejected due to calculating confidence interval in the fitness function. If the number of rejected instance increases, the size of the data will be increases. Analyzing and reducing rejected instances reaches acceptable level, then the online tests will be performed.

REFERENCES

[1] D.E. Denning, "An intrusion-detection model," Software Engineering, IEEE Transactions on, pp. 222-232, 1987.

[2] C.Kruegel and T. Toth, "A survey on intrusion detection systems," in TU Vienna, Austria, 2000.

[3] J.M. Estevez-Tapiador, P. Garcia-Teodoro, and J. E. Diaz-Verdejo, "Anomaly Detection Methods in Wired Networks: A Survey and Taxonomy," Computer Communications, vol. 27, pp. 1569-1584, 2004.

[4] S.H. Amer and J. Hamilton, "Intrusion Detection Systems (IDS) Taxonomy-A Short Review," Defense Cyber Security, vol. 13, 2010.

[5] A.Lazarevic, L. Ertöz, V. Kumar, A. Ozgur, and J. Srivastava, "A Comparative Study of Anomaly Detection Schemes in Network Intrusion Detection," in SDM, 2003, pp. 25-36.

[6] V.Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Computing Surveys (CSUR), vol. 41, p. 15, 2009.

[7] G.Münz, S. Li, and G. Carle, "Traffic anomaly detection using kmeans clustering," in GI/ITG Workshop MMBnet, 2007.

[8] L.Kaufman and P. J. Rousseeuw, Finding groups in data: an introduction to cluster analysis vol. 344. New York: John Wiley & Sons, 1990.

[9] D.E. Golberg, "Genetic algorithms in search, optimization, and machine learning," Addion wesley, vol. 1989, 1989.

[10] W.Li, "Using genetic algorithm for network intrusion detection," Proceedings of the United States Department of Energy Cyber Security Group, pp. 1-8, 2004.

[11] P.Gupta and S. K. Shinde, "Genetic algorithm technique used to detect intrusion detection," in Advances in Computing and Information Technology, ed: Springer, 2011, pp. 122-131.

[12] P.G. Majeed and S. Kumar, "Genetic algorithms in intrusion detection systems: A survey," International Journal of Innovation and Applied Studies, vol. 5, pp. 233-240, 2014.

[13] C.Z. Janikow and Z. Michalewicz, "An experimental comparison of binary and floating point representations in genetic algorithms," in ICGA, 1991, pp. 31-36.

[14] C.C. Coello, G. B. Lamont, and D. A. Van Veldhuizen, Evolutionary algorithms for solving multi-objective problems: Springer Science & Business Media, 2007.

[15] S.Sivanandam and S. Deepa, Introduction to genetic algorithms: Springer Science & Business Media, 2008.

[16] Z.Michalewicz, Genetic algorithms+ data structures= evolution programs: Springer Science & Business Media, 1996.

[17] Y.Chen, Y. Li, X.-Q. Cheng, and L. Guo, "Survey and taxonomy of feature selection algorithms detection system," in Information Security and Cryptology, 2006, pp. 153-167.

[18] H.G. Kayacik, A.N. Zincir-Heywood, and M. I. Heywood, "Selecting features for intrusion detection: A feature relevance analysis on KDD 99 intrusion detection datasets," in Proceedings of the third annual conference on privacy, security and trust, 2005.

[19] A.I. Madbouly, A. M. Gody, and T. M. Barakat, "Relevant Feature Selection Model Using Data Mining for Intrusion Detection System," arXiv preprint arXiv:1403.7726, 2014.